Quantifying Significant Expansion of a Federal Data Center

October 20 2016. Version 2.2

Background

In compliance with the Federal Information Technology Acquisition Reform Act (FITARA) and in support of the government's shared services efforts, the General Services Administration (GSA) Office of Government-wide Policy (OGP) serves as the Managing Partner for the Data Center Optimization Initiative (DCOI).

As outlined in the Office of Management and Budget's (OMB)'s Memorandum M-16-19, bullet #9 on page 2:

> *"The General Services Administration (GSA) Office of Government-wide Policy (OGP) will coordinate with OMB to define thresholds for what constitutes **significant** expansion within 60 days of publication of this memorandum."*

OGP provided analysis pursuant to fulfilling the Federal CIO's directive and offered evidence-based recommendations to OMB on how to define the aforementioned threshold.

Approved Guideline

On September 28, 2016 OMB approved OGP's recommendation of the following threshold of ***above 18% or above 660 square feet*** at the individual Tiered Data Center[1] level. Expansions that fall farther than one standard deviation away from the mean should be considered statistically significant. Recommended indicators are based on finding a statistically significant threshold in the Integrated Data Collection (IDC)[2] data trends reported to OMB.

---

[1] OMB Memorandum M-16-19, "Data Center Optimization Initiative (DCOI), August 1, 2016, https://www.whitehouse.gov/sites/default/files/omb/memoranda/2016/m_16_19_1.pdf.

[2] OMB Memorandum M-13-09, "Fiscal Year 2013 PortfolioStat Guidance: Strengthening Federal IT Portfolio Management," March 27, 2013, https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-09.pdf.

OGP developed a model to perform calculations, which was used to derive the following guiding measures for the significant expansion threshold, as shown in Figure 1.

Tiered Data Center Level Significant Expansion Based on Gross Floor Area (GFA)			
Mean	(189)		The simple average of the numbers (the average GFA of all data centers at the Agency level).
Variance	12,496,973		The average of the squared differences from the Mean.
Standard Deviation	3,535		The square root of the Variance. Measures how spread out Agencies' GFA numbers are. Tells us how wide or concentrated GFA expansion data is. Interpreted as how far away the typical/average observation is from the mean.
GFA % Change	18.73%	$\frac{Standard\ Deviation}{Mean}$ $(\frac{3{,}535}{189})/100 = 18.73\%$	Expansion above ***18.73%*** of GFA may be considered significant expansion at the ***Tiered Data Center*** level.
Relative Standard Deviation (RSD)	662	$3{,}535 * 18.73\% = 662\ Sq.Ft.$	Expansion above ***662*** GFA may be considered significant expansion at the ***Tiered Data Center*** level.

Figure 1

Methodology

In the absence of having accurate year-by-year square footage data for individual federal data centers, OGP obtained the IDC Q1 data for Gross Floor Area (GFA) for the following years: 2014, 2015, and 2016. OGP aggregated the data obtained from IDC at the Government-wide (Agency), Bureau (Sub-Agency or Component), and Data Center levels to calculate measures in order to define a reasonable threshold for what constitutes significant expansion.

Measure Definitions:

- The **Mean** is the average expansion expressed in square footage.
- The **Variance** is the average of the squared differences from the Mean expressed in square footage.
- The **Standard Deviation** is the square root of the Variance expressed in square footage. Used to quantify the amount of variation or dispersion of a set of data values.

- The **GFA % Change** is calculated using [(Standard Deviation / Mean) / 100], which provides a statistical threshold for significant expansion expressed as a percentage.

- The **Relative Standard Deviation (RSD)** provides a statistical threshold for significant expansion expressed in square footage. It is the absolute value of the coefficient of variation describing the spread of data with respect to the mean.

Based on the IDC data sets reported to OMB for Q1 of 2014, 2015, and 2016, the model was used to calculate the threshold of statistical significance reflecting the relative standard deviation of changes in data center size. The results indicated 660 sq. ft. as a threshold marking any data point beyond that as significant expansion for Tiered data centers.

The model performed the following calculations:

- Difference between Q1 2014 and Q1 2015 in terms of expansions and contractions.
- Difference between Q1 2015 and Q1 2016 in terms of expansions and contractions.
- Average all expansions and contractions across three-years period for each of Q1 for 2014, 2015, and 2016 years. This step accounts for all the expansions and contractions between Q1 2014 to Q1 2015 to Q1 2016 to smooth out all the data sets removing any statistical bias in order to avoid penalizing or rewarding data centers for expanding or contracting.
- Find the Mean, Variance, Standard Deviation, and Relative Standard Deviation based on square footage and percent deviation from the mean in order to find a point of a statistical significant threshold based on the entire data population from Q1 2014 to Q1 2016 after smoothing out expansions and contractions.

A radar chart as shown in Figure 3 was used instead of a bell-shaped curve that displays the normal distribution of the data to visually display and call out outliers.

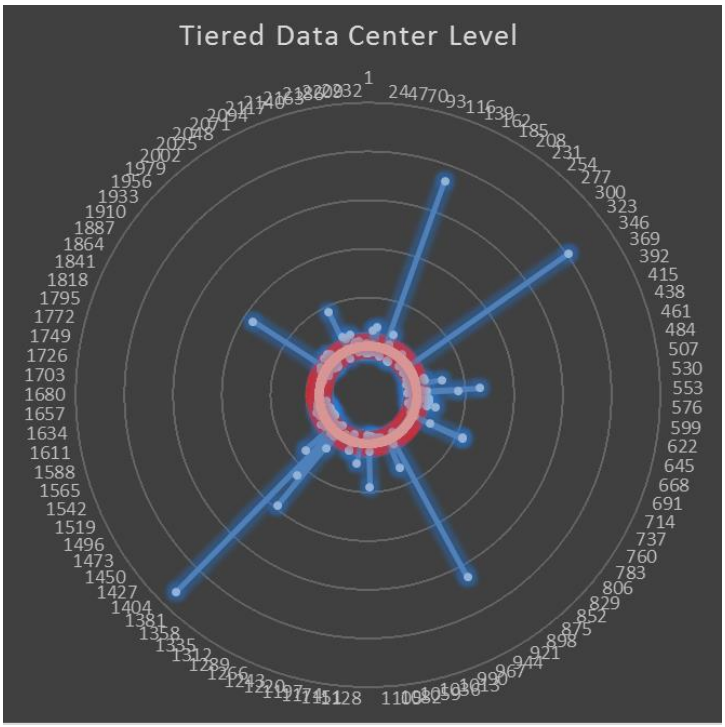

Figure 1

Disclaimer

To determine significant expansion based on data center square footage has its flaws. Square footage on its own may not be an indicator of significant expansion as it does not capture the cost associated with data center expansion through density (i.e. processing powering and storage capability increasing rapidly, virtualization of servers, mezzanines of rack structures, and high density server racks). Significant expansion of data center capacity can occur through a shift to modern equipment and methods without any expansion of the building footprint or square footage.

The IDC data is reflected on Q1 for each of the 2014, 2015, and 2016 years. Many data quality issues were found in the data sets to include data gaps and defects that may limit the findings or slightly affects the results of the significant expansion model.